IR system and features: study of Indian scenario

Dr. Sarika Sawant (sarikashpt@gmail.com), Assistant Professor, SHPT School of Library Science, SNDT Women's University, Mumbai, INDIA

ABSTRACT

Purpose: The main goal was to study issues concerning IR software/system involved in developing and managing an institutional repository.

Methodology: For the present study survey research method was found to be most suitable. The data collection tool applied for the study was web questionnaire, which was created with the help of software provided by surveymonkey.com.

Findings: It was observed that 79% institutions had implemented DSpace Institutional Repository software package. Respondents assigned top-rank to IR-system feature - End-user interface. It was found that all Institutional Repositories supported Text (HTML, Postscript, PDF, Spreadsheet etc) file formats. Half of the respondents marked bitstream copying as a long-term preservation strategy. Almost all institutional repositories were OAI-PMH compliant.

Limitations: Only Indian institutional repositories were studied and the findings were compared with other studies.

Originality: In detail IR system/software used in developing repositories were studied which is one of the first study of its kind. The present study discovered existence of 16 functional IRs some of which were not registered in any of the directories such as ROAR, Open DOAR.

Keywords: Institutional repositories, IR system, IR features

Introduction

Indian Scenario

In India, there are about 16 functional institutional repositories developed by research institutions those are of national and international importance such as Indian Institute of Science, Indian Institute of Management etc. Apart from institutional repositories, Subject specific repositories also exist that store and provide access to subject specific collections of documents. These repositories accept scholarly publications from any professional or researcher who belongs to the respective subject. *Librarian's Digital Library* (LDL) of Documentation Research and Training Centre (DRTC), Bangalore is an example of subject-specific repository for the library and information professionals. Another subject-specific repository established in India is *OpenMed@NIC*, maintained by National Informatics Centre, New Delhi. OpenMed@NIC stores and provides access to biomedical literature. Other kind of digital repositories existing in India stores and provides access to document type specific collections. Vidyanidhi of University of Mysore is an example of document type specific collection that stores and provides access to theses and dissertations (Cross institutional ETD repository). Vidyanidhi accepts any thesis or dissertation that has been accepted in any of the Indian universities or institutions (Fernandez, 2006).

IR Software's

The leading IR software packages, DSpace (MIT) and EPrints (Southampton) as well as Greenstone are available free under open source licenses, and there are at least half a dozen other possible packages. In theory, commercial document management or knowledge management software packages might also be suitable but are unlikely to be adopted given their costs.

According to Lynch (2006) making a decision can be complex and involves careful thought about factors such as what the repository will contain, how it will be used, the features that are wanted, and the local technical environment. It is also important to select

software with the right features. However, even the 'best' software may not have every feature an institution wants.

2 Objectives and Methods

The main goal was to study issues concerning IR software/software involved in developing and managing an institutional repository.

There were four broad objectives, which are as follows:

1. To know which software were used to develop institutional repositories and its features
2. To identify which interoperability standards and long-term preservation techniques were applied
3. To know which types of file formats IR supports
4. To know whether they were willing to migrate to new IR system in future

One of the first steps in the data gathering process was the identification of population i.e. all institutional repositories in India. To compile the list of institutional repositories the researcher used various sources of information such as: Literature; Search by search engines especially Google; Directories of archives / repositories; Blogs; Open Source Software websites etc. which resulted into identification of 16 institutional repositories.

To operationalise the study survey method was found to be most suitable. The data collection tool applied for the study was web questionnaire, which was created with the help of software provided by surveymonkey.com. After identification of institutional repositories (Table No. 1) and e-mail addresses of web administrators of these repositories, researcher sent e-mails containing URL of the web questionnaire and requested to fill the required data in the questionnaire.

Total 14 responses out of 16 were received making total response rate of 87.5% over the period of four months.

Table No. 1 List of institutional repositories considered for the study

Sr. No	Name of the IR	URL of the IR
1	Delhi University, New Delhi	http://eprints.du.ac.in/
2	ICFAI Business School, Ahmedabad	http://202.131.96.59:8080/dspace/
3	IIT Bombay (GR), Mumbai	http://dspace.library.iitb.ac.in/dspace/
4	IIT Bombay (ETD), Mumbai	http://www.library.iitb.ac.in/~mnj/gsdl/cgi-bin/library
5	Indian Institute of Astrophysics, Bangalore	http://prints.iiap.res.in/
6	Indian Institute of Management, Kozhikode	http://dspace.iimk.ac.in/
7	Indian Institute of Science (GR), Bangalore	http://eprints.iisc.ernet.in/index.html
8	Indian Institute of Science (ETD), Bangalore	http://etd.ncsi.iisc.ernet.in/
9	IIT Delhi, New Delhi	http://eprint.iitd.ac.in/dspace/
10	Indian Statistical Institute, Bangalore	http://library.isibang.ac.in:8080/dspace/
11	Indira Gandhi Institute of Development Research, Mumbai	http://202.54.18.153:8888/dspace/index.jsp
12	National Aerospace Laboratories, Bangalore	http://nal-ir.nal.res.in/
13	National Chemical laboratory, Pune	http://dspace.ncl.res.in/dspace/index.jsp
14	National Institute Of Oceanography, Goa	http://drs.nio.org/drs/index.jsp
15	National Institute of Technology, Rourkela	http://dspace.nitrkl.ac.in/dspace
16	Raman Research Institute, Bangalore	http://dspace.rri.res.in/

3 Results

3.1 Type of IR Software / System

Question was asked to respondents to choose the Institutional Repository software / system that they had pilot tested and implemented for developing institutional repository. The question was close ended and listed three options.

It was observed that 79% institutions i.e. 11 institutions had implemented DSpace Institutional Repository software package. This was not unexpected. DSpace was one of the first software packages specifically developed for IR services. Out of these 11 Institutional Repositories, 7 (64%) Institutional Repositories had first pilot tested DSpace

and then implemented it. The GNU Eprints and Greenstone were used by two (14.28%) and one (7.14%) institutions respectively. The data is presented in Table No. 2.

Table No. 2: IR Software / System

IR Software/System	Pilot Tested	Implemented	Percentage
DSpace	**7**	**11**	**79**
GNU Eprints	2	2	14.28
Greenstone (GSDL)	3	1	7.14

Additional comments were given by two respondents (14.29%). The respondent from RRI had commented that they had tested Greenstone in 2004. The version did not support input of documents from any computer. On the other hand in DSpace any legitimate user could register and be authorised to submit from any computer. Also DSpace had been used by different people to build a repository. Further the respondent was unsure whether the Greenstone metadata was OAI-PMH compliant or not while DSpace was.

The respondent from IGIDR had commented that GSDL did not provide workflow hence was not considered for IR implementation. Therefore both the institutions had opted for DSpace after testing GSDL.

Open Access Repositories in New Zealand project (OARINZ) was undertaken by a collaboration of Tertiary Institutions. As a part of this project, technical evaluation of Open Source Repository system/s was carried out for further enhancement and large-scale deployment across New Zealand. The findings suggested that DSpace could be accommodated within the national network because DSpace scored well in the overall evaluation (Open Access Repositories in New Zealand project, 2006).

A number of studies have indicated the usefulness of DSpace. Lynch and Lippincott (2005) surveyed academic member institutions of Coalition for Networked Information (CNI) to examine the current state of institutional repositories (IRs) in the United States. The findings had suggested that higher number of respondents i.e. 58% of respondents indicated that they were using DSpace. Another study done by Markey et al. (2007) had found that 46.4% respondents had developed IR using Dspace, out of which 19% pilot

had tested before implementation. The study clearly indicated that GSDL and EPrints were not the favourites.

Similar results were found in the study done by Bailey et al. (2006) who had observed that DSpace was most commonly used system by 23 (70%) out of 33 respondents, out of 23, 20 used it exclusively and 3 used it in combination with other systems. Two of the implementers used it in conjunction with CONTENTdm (commercial software) and the third respondents mentioned that they used DSpace in conjunction with ETD-db and Open Conference Systems (both open source software). Of the respondents that did not use DSpace, one respondent mentioned that they used open source Archimède software and two used commercial CONTENTdm software.

Further, Bailey reported that 28% of respondents had made no IR software modifications to enhance its functionality, 22% had made frequent changes to do so and 20% had made major modifications of the software.

3.2 IR-system Features

Respondents were asked in the next question to rate Institutional Repository systems, with regard to various capabilities such as technical issues and documentation. The question was close ended listing 15 options. Respondents were requested to choose more than one option, which were applicable to them.

The responses were quantified using a Likert type scale: VERY IMPORTANT 4 to LEAST IMPORTANT 1. (Very Important = 4; Important =3; Somewhat Important = 2; Least Important = 1). Apart from these options, two more options were given namely Don't Know and Non Applicable. Thus for each activity score were obtained. The scores were used to rank the activities, which is presented in Table No. 3.

Table No. 3: IR-system Features

IR-system Features	Score	Rank
End-user interface	**53**	**1**
Browsing, searching, and retrieving digital content	49	2
Adherence to open access standards	48	3

Supported file formats	48	3
Digital preservation	45	5
Formulating metadata for digital documents	45	5
Technical documentation	40	7
User authentication	40	7
Technical support	38	9
Authority control	37	10
Customization	36	11
Controlled vocabulary searching	34	12
Scalability (System growth and enhancement)	34	12
Multilingual support	27	14
Extensibility (Access to other campus systems and data)	26	15

Respondents assigned top rank to IR-system feature - End-user interface (score 53). This was followed by browsing, searching, and retrieving digital content (score 49). Supported file formats (score 48) and Adherence to open access standards (score 48) had equal rank. Multilingual support (score 27) and Extensibility (Access to other campus systems and data) (score 26) scored the lowest.

However, in the study done by Markey et al. (2007) it was found that the two top ranked IR-system features were - 'Supported file formats' and 'Adherence to open-access standard'. These two features ranked at 3rd positions in the present study.

3.3 File Formats

It was observed that all Institutional Repositories supported Text (HTML, Postscript, PDF, Spreadsheet etc) file formats. About 78.6% (11) respondents supported Image (TIFF, GIF, JPEG etc.) file formats. Equal number of respondents i.e. 57.14% (8) supported Audio (WAV, MP3 etc) and Video (MPEG, AVI etc) file formats. The data is presented in Table No. 4. There were three institutional repositories (21.43%) that supported all file formats. These were IIAP, IIMK and IITB (GR).

In addition to the listed file formats the respondent from IITB (ETD) mentioned in 'Others' that they supported one more file format i.e. Open document Text format.

Table No. 4: File Formats

File Formats	Number of Responses	Percentage

Text (HTML, Postscript, PDF, Spreadsheet etc)	**14**	**100**
Images (TIFF, GIF, JPEG etc)	11	78.57
Audio (WAV, MP3 etc)	8	57.14
Video (MPEG, AVI etc)	8	57.14
Datasets	6	42.86
Computer programs	3	21.43
CAD/CAM	3	21.43
Databases	3	21.43
Others	1	7.14

However, in the study done by Bailey et al. (2006) had observed that 74% of respondents (out of 37) indicated that they accept any digital file type into the IR, but relatively few (26%) were committed to functional preservation of every file type. About 18% of respondents accepted and preserved specified file types. A few accepted certain file types but did not preserve them. Several respondents mentioned following the support levels outlined in MIT's DSpace guidelines (http://www.dspace.org/implement/policy-issues.html#digformats), which include full support and preservation for common file types such as PDF, XML, AIFF for audio, and GIF, JPEG, and TIFF for images, among others.

3.4 Long-term Preservation Strategies

Half of the respondents (50% i.e. 7 respondents) marked bitstream copying as a long-term preservation strategy. About 21.40% (3) of the respondents marked: Durable, Persistent Media (where you preserve the physical media, or CD, on which object is stored). Standards as well as Preservation metadata also had the same score (21.40% i.e. 3 respondents) as above. Emulation as a preservation strategy was not exercised by any of the institutional repository. There were four (28.60%) institutional repositories namely ICFAI, NAL, NIO and NITR who were not following any long-term preservation strategy. The data is presented in Table No. 5.

Table No. 5: Long-term Preservation Strategies

Long-term Preservation Strategies	Number of Responses	Percentage
Bitstream Copying	**7**	**50.00**
Durable, Persistent Media (where you preserve the physical media, or CD, on which object is stored)	3	21.40

Migration	2	14.30
Standards	3	21.40
Emulation		
Encapsulation	2	14.30
Preservation metadata	3	21.40
No long term preservation strategy is employed	4	28.60

In the study done by Bailey et al. (2006) it was found that out of those who accept any file type 47% preserve specified file types using data migration and other techniques. The next most common arrangement (26%) was to accept and preserve any file type. However in the present study 14.30% (2) of respondents employed migration as a long-term preservation strategy.

3.5 Interoperability Standards

About 92.86% (13) institutional repositories were OAI-PMH compliant. Only one (7.14%) institutional repository supported Metadata Encoding and Transmission Standard (METS) and 14.28% (2) Institutional Repositories were OpenURL compliant. The data is presented in Table No. 6.

Table No. 6: Interoperability Standards

Interoperability Standards	Number of Responses	Percentage
IR supports OAI-PMH	13	92.86
Metadata Encoding and Transmission Standard (METS)	1	7.14
IR is OpenURL compliant	2	14.28
Others	1	7.14

It was interesting to note that IIMK institutional repository supported all three interoperability standards. The respondent from ICFAI had mentioned in 'Others' that their IR support OAI-PMH but had not yet been enabled for OAI access.

Similar results were found in the study done by Bailey et al. (2006). The survey findings reported that 94% of respondents (i.e. Out of 37 respondents) had indicated that their IR supports the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), 54% (i.e. Out of 37 respondents) had ensured that their systems are OpenURL compliant.

3.6 Migration to New Institutional Repository Software / System

The last question was asked to the respondents about their plans for migration to new Institutional Repository software / system. The question was kept open ended because it was thought that there could be different opinions of respondents to this question.

About 57.14% (8) of respondents mentioned that they did not have any plans to migrate to new IR software / system at present. About 35.71% (5) respondents skipped the question. Only one respondent (IITB (ETD) 7.14%) had mentioned that they have a plan to migrate to new IR software / system. The data is presented in Table No. 7.

Table No. 7: Migration Plans

Migration Plans	Number of Responses	Percentage
No plans for migration	8	57.14
Plans for migration	1	7.14

In the study done by Markey et al. (2007) it was found that 56% of respondents thought they would migrate to new IR software within the next three years. About 40% thought they would migrate in the next four to six years. The remaining 4% said that they would continue with their present system for seven or more years.

IITB had 2 IRs one general repository in DSpace which was implemented in 2007 and one electronic theses and dissertation repository in Greenstone which was implemented in 2003. The respondent had mentioned that they have a plan to migrate from Greenstone to DSpace for their electronic theses and dissertation repository. They must had thought of DSpace being more suitable software for development of IR. So they thought of migrating from Greenstone to DSapce.

4 Conclusion

It was observed that 79% (11) institutions had implemented DSpace Institutional Repository software package. Out of 79% (11) of institutions, 64% Institutional Repositories had first pilot tested DSpace and then implemented it. Respondents assigned top-rank to IR-system feature - End-user interface (score 53). This was followed by browsing, searching, and retrieving digital content (score 49), Supported file formats

(score 48) and Adherence to open access standards (score 48) received high rank. It was found that all Institutional Repositories supported Text (HTML, Postscript, PDF, Spreadsheet etc) file formats. Half of the respondents (50% i.e. 7 respondents) marked bitstream copying as a long-term preservation strategy. About 92.86% (13) institutional repositories were OAI-PMH compliant. About 57.14% (8) of respondents mentioned that they did not have any plans to migrate to new IR.

REFERENCES

Bailey C W, Coombs K, Emery J et al. (2006) Institutional Repositories. *SPEC Kit 292*. (accessed 17 September 2009) http://www.arl.org/spec/SPEC292web.pdf

Harnad S, Swan A (2008) India, Open Access, the Law of Karma and the Golden Rule. *DESIDOC Bulletin of Information Technology* 28: 1 (accessed 12 May 2008) http://eprints.ecs.soton.ac.uk/14432/1/india.pdf

Fernandez L (2006) Open Access Initiatives in India - an Evaluation. (accessed 10May 2009) http://units.sla.org/division/dst/Annual%20Conference%20Contributed%20Papers/2006papers/Leila%20Fernandez%20open%20access%20India.pdf

Lynch C A (2006) Open Computation: Beyond Human-Reader-Centric Views of Scholarly Literatures. In: Jacobs N (ed) *Open Access: Key Strategic, Technical and Economic Aspects*. Oxford: Chandos Publishing, 185-193

Markey K, Rieh S Y, Jean B S et al. (2007) Census of Institutional Repositories in the United States: MIRACLE Project Research Findings. (accessed 11 January 2009) http://www.clir.org/pubs/reports/pub140/pub140.pdf

Open Access Repositories in New Zealand (OARINZ) project (2006) Technical Evaluation of Selected Open Access Repositories in New Zealand. (accessed 16 September 2009) https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf